International Journal of Electronics and
Communication Engineering (IJECE)
ISSN(P): 2278-9901; ISSN(E): 2278-991X
Vol. 4, Issue 1, Jan 2015, 21-28
© IASET

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

DETECTING HIDDEN DATA USING FIRST AND SECOND ORDER STATISTICS

MONISHA SHARMA[1] & SWAGOTA BERA[2]

[1]Professor, Department of Electronics & Tele, SSCET, Bhilai, India

[2]Associate Professor, Department of Electronics & Tele, SSIET, Durg, India

ABSTRACT

Steganalysis is the art of detecting the presence of hidden data in files. Universal steganalysis is general class of steganalysis techniques which can be implemented with any steganographic embedding algorithm, even an unknown algorithm. In this paper, the detection method includes the variation in the feature vectors corresponding to cover and stegoimages. The features are calculated as an L_1 norm of the difference between a specific macroscopic functional. The functionals are built from marginal and joint statistics of DCT coefficients. The features are calculated directly from DCT co-efficients. So, the detection is possible for the JPEG images. Three different steganographic paradigms are tested and compared.

KEYWORDS: Steganography, Steganalysis, Cover Image, Stego image, Cover Image, Attack, Least Significant Bit (LSB), DCT

1 INTRODUCTION

Reported in USA TODAY AP by Jack Kelley in 2002 that hidden in the X-rated pictures on several Web sites and the posted comments on sports chat rooms may lay the encrypted blueprints of the next terrorist attack against the United States or its allies. It sounds farfetched, but U.S. officials and experts say it's the latest method of communication being used by Osama bin Laden and his associates to outfox law enforcement. Bin Laden, indicted in the bombing in 1998 of two U.S. embassies in East Africa, and others are hiding maps and photographs of terrorist targets and posting instructions for terrorist activities on sports chat rooms, pornographic bulletin boards and other Websites, U.S. and foreign officials say.

Reported in German newspaper by Sean Gallagher in 2012 that, digital steganography hides files in plain sight, concealed in image and media files. When a suspected al-Qaeda member was arrested in Berlin in May of 2011, he was found with a Memory card with a password-protected folder and the files with in it were hidden. But, as the, computer forensics experts from the German Federal Criminal Police (BKA) claim to have eventually uncovered its contents—what appeared to be a pornographic video called "KickAss." Within that video, they discovered 141 separate text files, containing what officials claim are documents detailing al-Qaeda operations and plans for future operations—among them, three entitled "Future Works," "Lessons Learned," and "Report on Operations."

The above facts concludes that the hiding technique is becomes a vital tool for resulting the disaster from 1998 to till date. So it is important to work in improving the detection technique. Various approaches are discussed by the different researchers in the area of steganalysis. Broadly, there are two approaches to the problem of steganalysis, and one is to come up with a steganalysis method specific to a particular steganographic algorithm known as embedding algorithm based steganalysis techniques. For decoding we must know the encoding algorithm. Complete recovery of the secret data is possible. The other technique is more general class of steganalysis techniques pioneered independently can be designed to work with any steganographic embedding algorithm, even an unknown algorithm. Such techniques have subsequently been called universal steganalysis techniques or blind steganalysis techniques. A steganalysis technique specific to an embedding method would give very good results when tested only on that embedding method, and might fail on all other steganographic algorithms. But blind steganalysis is the universal one which can be implemented to any stegoimage. Featuresof typical natural images which can get violated when an image undergoes some embedding process. Hence, designing a feature classification based universal steganalysis technique consists of tackling two independent problems. The first is to find and calculate features which are able to capture statistical changes introduced in the image after the embedding process. The second is coming up with a strong classification algorithm which is able to maximize the distinction captured by the features and achieve high classification accuracy. Typically, a good feature should be accurate, consistent and monotonic in capturing statistical signatures left by the embedding process. Prediction accuracy can be interpreted as the ability of the measure to detect the presence of a hidden message with minimum error on average. Similarly, prediction monotonicity signifies that the features should ideally be monotonic in their relationship to the embedded message size. Finally, prediction consistency relates to the feature’s ability to provide consistently accurate predictions for a large set of steganography techniques and image types. This implies that the feature should be independent on the type and variety of images supplied to it .Embedding techniques affect different aspects of images.

2 LITERATURE SURVEY

Survey of some latest research papers in the present field is done[8,11]. The research in this field is started from 1995. A brief introduction is given about the papers .The research paper in which the steganalysis techniques is related to the work in this paper is only discussed.

Steganalysis can also be classified based on the detection parameters .If the detection is done in the basis of the differences in the visual texture of the stego image and cover image then known as visual attacking whereas if the detection is done in the basis of the variation in the statistical parameters of the stego image and cover image the known as the statistical attacking. These attacking technique need the information of cover image and stego image. The universal steganalysis is better technique in which the technique can be implemented to any randomly incoming images and detection can be done. Any portion of the incoming image may alter after applying the detection algorithm. If after removing the noise effect the alteration remains then it is detected that the secret data is hidden in it. If the detection result indicate the presence of hidden information , then the analytic will try to recover the secret data by applying the decoding algorithm in the hit and trial method. Prediction accuracy can be interpreted as the ability of the measure to detect the presence of a hidden message with minimum error on average. The feature should be independent on the type and variety of images supplied to it [4,5,7].

Different approaches are given by the various researchers for feature extraction from images. The authors argue that most of the specific steganalysis techniques concentrate on first order statistics, i.e. histogram of DCT coefficients. Quadratic mirror filters (QMF) are used to decompose the image, after which higher order statistics such as mean, variance, kurtosis, and skewness are calculated for each subband. Also the error obtained from an

optimal linear predictor of coefficient magnitudes of each sub band is used as a second set of features. The image features can be calculated by Markovs transition statistics, DCT functionals, Moments of characteristic function using wavelet decomposition, statistical moments. In all the above methods, the calculated features are used to train a classifier, which in turn is used to classify clean and stego images. Different classifiers have been employed by different authors; Ascribes uses a MMSE Linear predictor, whereas Farad uses a Fisher linear discriminate and also a Support Vector Machine (SVM) classifier. SVM classifiers seem to have much better performance in terms of classification accuracy compared to linear classifiers since they are able to classify non-linearly separable features [8, 11].

The steganography and steganalysis has already discussed in the previous survey report [8]. Few Image Statistical and visual steganalysis techniques are implemented and discussed in previous paper [9, 10].

3 METHODOLOGY

3.1 First Order Features

The statistical features are calculated from the DCT coefficient .The simplest first order statistic of DCT coefficients is the histogram. Suppose, $d_k(i, j)$ is the DCT coefficient array with quantized value . Q(i, j), i, j = 1,…,8, k = 1, …, B represents the quantized value of the JPEG file. The symbol $d_k(i, j)$ denotes the (i, j)-th quantized DCT coefficient in the k-th block (there are total of B blocks). The global histogram of all 64k DCT coefficients will be denoted as H_r, where r = L, …, R, L = $\min_{k,i,j} d_k(i, j)$ and R = $\max_{k,i,j} d_k(i, j)$. Many of the steganographic programs preserves the global histogram but fails to preserve the histogram of the individual DCT modes. Thus, we add individual histograms for low frequency DCT modes to our set of functionals. For a fixed DCT mode (i, j), let h_r^{ij} , r =L, …, R, denote the individual histogram of values $d_k(i, j)$, k = 1, …, B. We only use histograms of low frequency DCT coefficients because histograms of coefficients from medium and higher frequencies are usually statistically unimportant due to the small number of non-zero coefficients. For a fixed coefficient value d, the dual histogram is an 8×8 matrix g_{ij}^{d} where δ (u,v)=1 if u=v and 0 otherwise. In words, g_{ij}^{d} is the number of how many times the value d occurs as the (i, j)-th DCT coefficient over all B blocks in the JPEG image. The dual histogram captures how a given coefficient value d is distributed among different DCT modes.

$$g_{ij}^{d} = \sum_{k=1}^{B} \delta\left(d, d_k(i,j)\right)$$

3.2 Second Order Features

The natural images can exhibit higher-order correlations over distances larger than 8 pixels, individual DCT modes from neighboring blocks are not independent. Thus, the features that capture inter-block dependencies can be violated by the various steganographic algorithms. Let I_r and I_c denote the vectors of block indices while scanning the image "by rows" and "by columns", respectively. The first functional capturing inter-block de-pendency is the "variation" V defined as

$$V = \frac{\sum_{i,j=1}^{8}\sum_{k=1}^{|I_r|-1}\left|d_{I_r(k)}(i,j)-d_{I_r(k+1)}(i,j)\right|+\sum_{i,j=1}^{8}\sum_{k=1}^{|I_c|-1}\left|d_{I_c(k)}(i,j)-d_{I_c(k+1)}(i,j)\right|}{|I_r|+|I_c|}$$

Most steganographic techniques in some sense add entropy to the array of quantized DCT coefficients and thus are more likely to increase the variation V than decrease. Embedding changes are also likely to increase the discontinuities along the 8×8 block boundaries. In fact, this property has proved very useful in steganalysis in the past [6,10,12]. Thus, we include two blockiness measures B_α, α = 1, 2, to our set of functionals. The blockiness is calculated from the decompressed JPEG image and thus represents an "integral measure" of inter-block dependency over all DCT modes over the whole image:

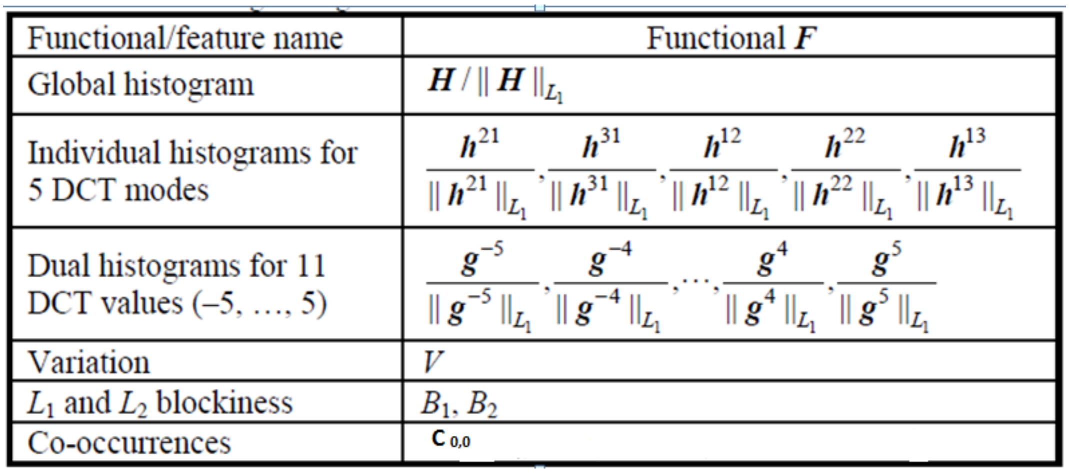
$$B_{\alpha} = \frac{\sum_{i=1}^{[(M-1)/8]}\sum_{j=1}^{N}\left|x_{8i,j}-x_{8i+1,j}\right|^{\propto}+\sum_{j=1}^{[(N-1)/8]}\sum_{i=1}^{M}\left|x_{i,8j}-x_{i,8j+1}\right|^{\propto}}{N[(M-1)/8]+M[(N-1)/8]}$$

In the expression above, M and N are image dimensions and $x_{ij}$are grayscale values of the decompressed JPEG image. The final three functionals are calculated from the co-occurrence matrix of neighboring DCT coefficients. Recalling the notation, L ≤ d_k(i, j) ≤ R, the co-occurrence matrix **C** is a square D×D matrix, D = R – L + 1, defined as follows has a sharp peak at (0,0) and then quickly falls off.

$$C_{st} = \frac{\sum_{k=1}^{|I_r|-1}\sum_{i,j=1}^{8}\delta\left(s,d_{I_r(k)}(i,j)\right)\delta\left(t,d_{I_r(k+1)}(i,j)\right)+\sum_{k=1}^{|I_c|-1}\sum_{i,j=1}^{8}\delta\left(s,d_{I_c(k)}(i,j)\right)\delta\left(t,d_{I_c(k+1)}(i,j)\right)}{|I_r|+|I_c|}$$

Due to the approximate symmetry of C_{st}around (s, t) = (0, 0), the C_{st} for (s, t)∈{(0,1), (1,0), (–1,0), (0,–1)} are strongly positively correlated in an image. The same is true for the group (s, t)∈{(1,1), (–1,1), (1,–1), (–1,–1)}.

Table 1: List of the Functionals



Functional/feature name	Functional F
Global histogram	$\boldsymbol{H} / \Vert \boldsymbol{H} \Vert_{L_1}$
Individual histograms for 5 DCT modes	$\frac{\boldsymbol{h}^{21}}{\Vert \boldsymbol{h}^{21} \Vert_{L_1}}, \frac{\boldsymbol{h}^{31}}{\Vert \boldsymbol{h}^{31} \Vert_{L_1}}, \frac{\boldsymbol{h}^{12}}{\Vert \boldsymbol{h}^{12} \Vert_{L_1}}, \frac{\boldsymbol{h}^{22}}{\Vert \boldsymbol{h}^{22} \Vert_{L_1}}, \frac{\boldsymbol{h}^{13}}{\Vert \boldsymbol{h}^{13} \Vert_{L_1}}$
Dual histograms for 11 DCT values (–5, …, 5)	$\frac{\boldsymbol{g}^{-5}}{\Vert \boldsymbol{g}^{-5} \Vert_{L_1}}, \frac{\boldsymbol{g}^{-4}}{\Vert \boldsymbol{g}^{-4} \Vert_{L_1}}, \ldots, \frac{\boldsymbol{g}^{4}}{\Vert \boldsymbol{g}^{4} \Vert_{L_1}}, \frac{\boldsymbol{g}^{5}}{\Vert \boldsymbol{g}^{5} \Vert_{L_1}}$
Variation	V
L_1 and L_2 blockiness	B_1, B_2
Co-occurrences	$C_{0,0}$

4 EXPERIMENTS AND RESULTS

4.1 Image Set

An image set consisting of 100 JPEG images with quality factors ranging from 70 to 90 is used in our experimental work. Each image was cropped (central portion) to the dimension of either 640 X 480. Some sample images are given in Figure (1).

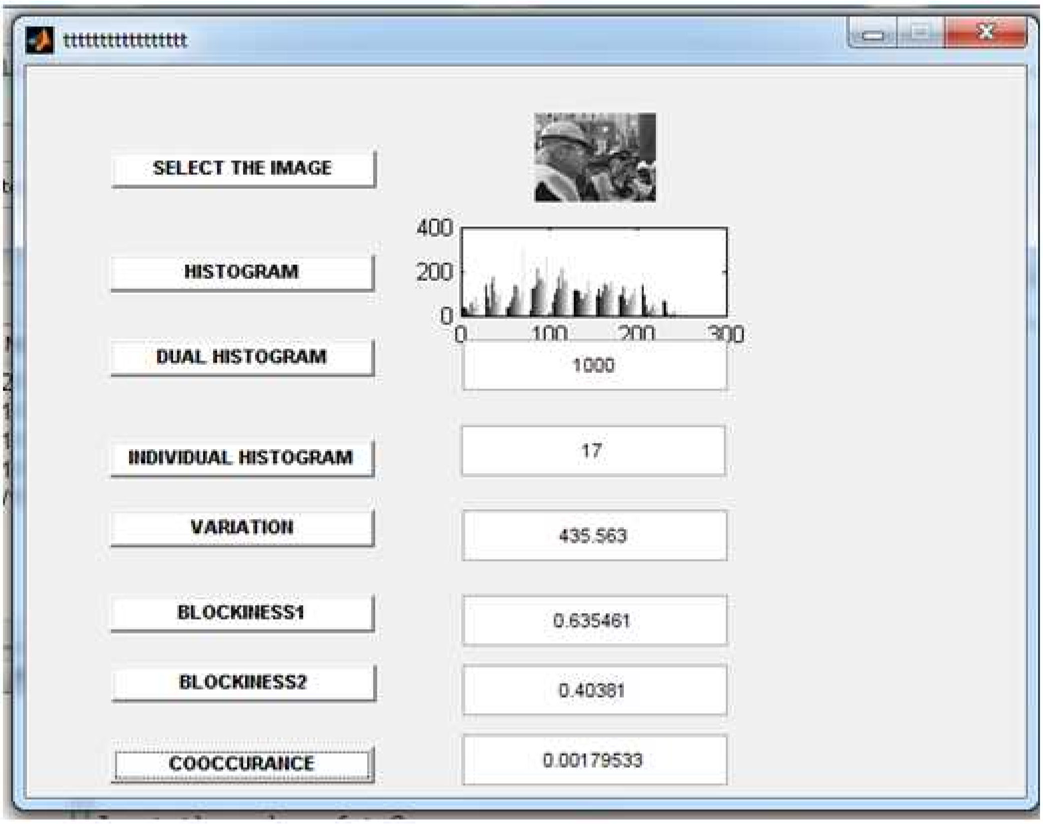
Figure 1: Some Sample Images Used in This Experimental Work

4.2 Stego Images Generation

On the basis of above approaches the steganalysis algorithm is designed using the MATLAB software and implemented to the stego image database, where database includes few different images of different size and formats encoded with JPEG Steganography technique for different capacties [14,15,16].

4.3 Experimental Results for First and Second Order Statistics

The variations in the statistical values is represented in the figure below after hiding the data in the image. The values of the functionals get varies if the capacity of the secret data get varies. In Figure (5) the functional values are represented for the stego images which is generated by the F5 JPEG steganographic technique and in Figure (6) the functional values are represented for the stego images which is generated by the Outguess JPEG steganographic technique. The same functionals are also calculated for the stego images generated by MB1, Jsteg, Wavelet based steganographictechniques. The stego images generated are of different capacities of 0.1,0.2 and 0.05 bpnc.



DUAL HISTOGRAM 1000

INDIVIDUAL HISTOGRAM 17

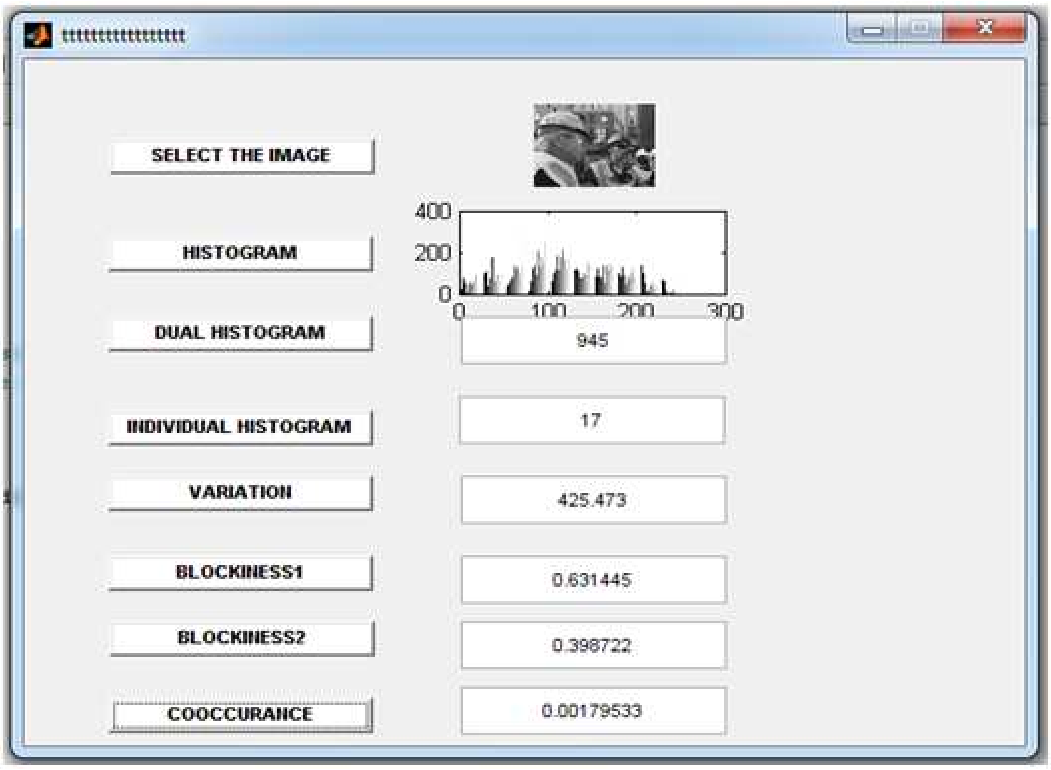
VARIATION 435.563

BLOCKINESS1 0.635461

BLOCKINESS2 0.40381

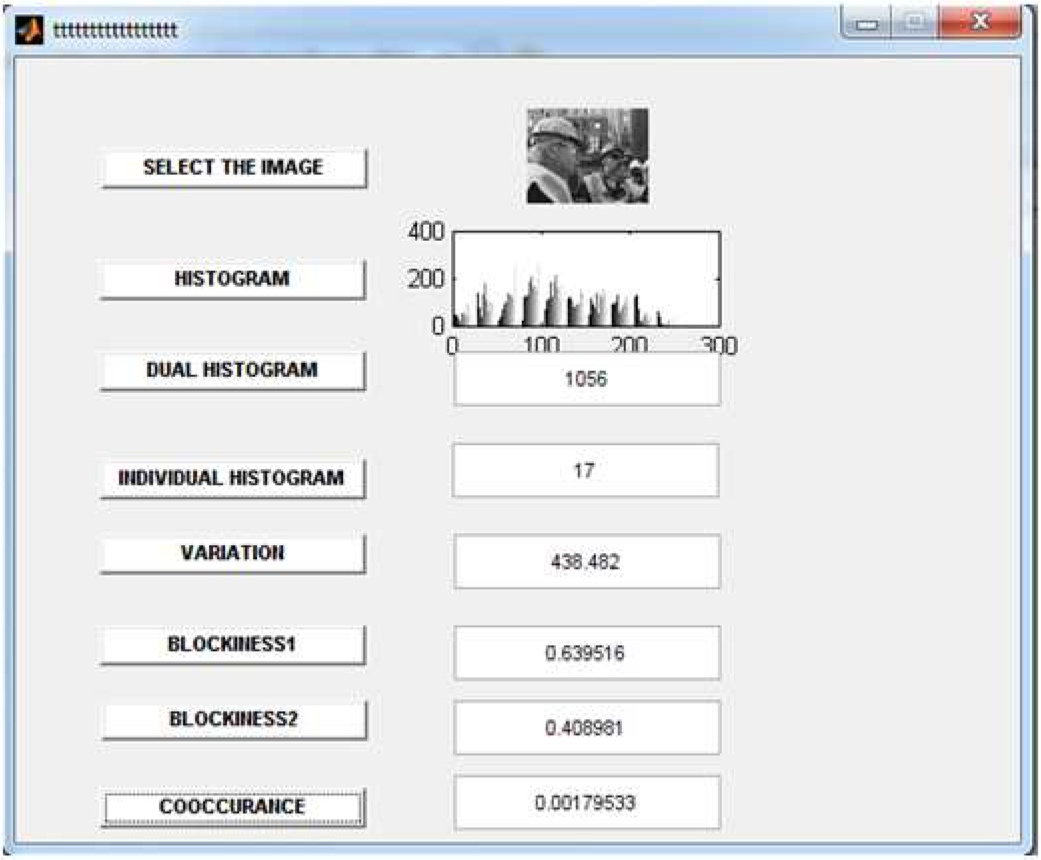
COOCCURANCE 0.00179533

Figure 2: Functional Values for the Cover Image



SELECT THE IMAGE	
HISTOGRAM	
DUAL HISTOGRAM	945
INDIVIDUAL HISTOGRAM	17
VARIATION	425.473
BLOCKINESS1	0.631445
BLOCKINESS2	0.398722
COOCCURANCE	0.00179533

Figure 3: Functional Values for the Stego Image Generated by F5 (bpnc 0.1)



SELECT THE IMAGE	
HISTOGRAM	
DUAL HISTOGRAM	1056
INDIVIDUAL HISTOGRAM	17
VARIATION	438.482
BLOCKINESS1	0.639516
BLOCKINESS2	0.408981
COOCCURANCE	0.00179533

Figure 4: Functional Values for the Stego Image Generated by Outguess (bpnc 0.1)

5 DISCUSSION AND CONCLUSIONS

The above obtained features will be helpful to train and design a classifier for the purpose of detecting the hidden data for an unknown image. This technique will be effective to detect the modern steganographic methods for JPEG images: OutGuess, F5, and MB1.

- There exist inter block dependencies within the image which is explore to calculate the statistics.
- Taking absolute decreases the calculation complexity.
- A scheme that preserves marginal statistics of DCT coefficients and the co-occurrence matrix (which captures block-to-block dependencies) is likely to exhibit improved resistance to attacks.

REFERENCES

1. Kelly, J. (2001). *Terror groups hide behind web encryption,* USA Today, 2 May 2001, Retrieved from http://www.usatoday. com/life/cyber/tech/2001- 02- 05-binladen.htm.

2. Anonymous, *what is steganography?.*Retrieved from www.tech-faq.com/steganography.html.

3. Maile, A., Zhanna, L., Ana, S., & Yasemin, Y. (2002, April). Image Encryption Using LSB/MSB. Term Project, CpE-462.

4. Anonymus. *Miscellaneous Steganographic.* Retrived fromscien.stanford.edu/class/psyh221/project /05/vvikram/stegomisc.htm.

5. Bernd. J. (II Ed.). (1993). Concepts, Algorithms, an Scientific Applications. *Digital Image processing.* Springer-Verlag.

6. Wolfgang, R.B., Delp, E.J. (1996). A watermark fordigital Image.(16-19 Sep1996). *International Conference on Images Processing,* Lausanne, Switzerland. IEEE, pp. 219–222.

7. Kharrazi, M., Sencar, T.H., & Memon, N., (2002). *Image Steganography: Concepts and Practices.* Department of Computer and Information Science Polytechnic University, Brooklyn, NY 11201, USA.

8. Bera, S. & Sharma, M. (2007). Survey on Steganographic Techniques & Steganalysis. (2007, 28-29Oct). *National Conference in Advances in electronics &Telecommunication Technologia.*vision-2020.

9. Bera, S. & Sharma, M. Steganalysis of Real Time Imageby Statistical Attacks.(2010). *International Journal ofEngineering Science and Technology,* Vol. 2(9), pp.4397-4406.

10. Bera, S. & Sharma, M. (2010). Steganalysis of theImage by Visual and Statistical Attack.i-manager's Journal of Electronics Engineering, ISSN-2229-7286.1(2):49-55.

11. Bera, S. & Sharma, M. (2012). A Review on blind stillimage Steganalysis Techniques Using Features Extractionand Pattern Classification Method. *International Journalof Computer Science, Engineering and Information Technology (IJCSEIT),* ISSN:2231-3117(Online);2231-3605(print). Vol. 2(3), pp.117-135.

12. Fridrich. J. 2005. *Feature basedsteganalysis for JPEG images and its implications for future design of steganographic schemes*. Information Hiding, Springer. 67–81.

13. Kumar M., (2012). *Steganography and Steganalysis of Joint Picture Expert Group (JPEG) Images.* Ph.D. Thesis, University of Florida.

14. Andrews Ker ,K. (2001). Retrieved fromhttp://www.outguess.org.

15. Latham, A. (2001). Retrived from http://wwwrn.inf.tudresden.de/~westfeld/f5.html.

16. A. Latham. (1999, August). *Jphide & seek.* [Online].Available: http://linux01.gwdg.de/latham/stego.html.